Niveau(x) d'enseignement: 3 Numéro EDUQ : 9670222
Centre d'intérêt: 640

Makdissi, Marie, 1940-

Présentation et résultats de l'utilisation en classe d'un logiciel d'évaluation des apprentissages / par Marie Makdissi. -- Sainte-Foy : INRS-Education, 1984.
10 f. ; 28 cm.

"Communication présentée au 4e colloque sur Les applications pédagogiques de l'ordinateur tenu à l'Université du Québec à Trois-Rivières le 18 octobre 1984". Code: C-129. Notes (part. bibliogr.) au bas des p.

Vedette(s)-secondaire(s): INRS-Education

Descripteur(s): Evaluation de logiciels * Evaluation de l'apprentissage * Evaluation du matériel didactique * Elaboration de tests * Enseignement assisté par ordinateur * Primaire (Ordre d'enseignement)

Le logiciel d'administration de tests, conçu et développé à l'INRS-Education en 1982-1983, a été évalué dans deux classes en 1983-1984. C'est un outil qui permet à l'enseignant de construire une banque à partir de laquelle sont produits des tests composés d'items choisis au hasard. Ces tests peuvent être imprimés ou présentés à l'écran; l'étudiant se sert alors du terminal pour répondre. Les trois mois d'observation-participation ont conduit à une évaluation positive du logiciel. Les deux enseignants qui ont expérimenté le logiciel ont reconnu sa valeur pour l'évaluation formative des étudiants et ont admiré sa souplesse pour produire des tests variés. Dans l'ensemble, les réactions des étudiants ont été très favorables. La version finale du logiciel, version améliorée, connue sous le nom de "Rétroaction", a été réalisée après l'évaluation. La Télé-Université l'a adoptée.

Type de document: r

EDUQ 8670222

Marie Makdissi

PRESENTATION ET RESULTATS DE L'UTILISATION EN CLASSE D'UN LOGICIEL D'EVALUATION DES APPRENTISSAGES.

INRS-Education, C-129, 1984.

EDUQ 8670222

Présentation et résultats de l'utilisation en classe d'un logiciel d'évaluation des apprentissages

par

Marie Makdissi 1940

Communication présentée au 4e colloque sur les applications pédagogiques de l'ordinateur tenu à l'Université du Québec à Trois-Rivières le 18 octobre 1984

C-129

INRS-Éducation
2383, Chemin Sainte-Foy
Sainte-Foy, Québec
G1V 1T1

Novembre 1984

DESCRIPTION DU LOGICIEL SOUMIS À L'ÉVALUATION

Globalement, le logiciel d'administration de tests, conçu et développé à l'INRS-Éducation au cours de l'année 1982-1983 et évalué en contexte scolaire pendant l'année 1983-1984, se veut un outil qui permette à l'enseignant(e) de construire une banque d'items à partir de laquelle sont générés des tests composés d'items choisis au hasard. Ces tests peuvent être imprimés ou encore, présentés à l'écran directement à l'étudiant qui se sert, à ce moment-là, du terminal pour y répondre.

D'une façon plus précise, le logiciel d'administration automatisée de tests possède quatre caractéristiques importantes pour lesquelles ont opté les membres de l'équipe de recherche chargée de sa conception.

La première particularité du système réside dans la possibilité offerte de générer un test basé sur un échantillon stratifié d'items. Cette stratification repose essentiellement sur un procédé de regroupement des items d'une banque, c'est-à-dire sur la formation de sous-ensembles d'items identifiés dans le logiciel par le mot "catégories".

Une deuxième particularité intéressante du logiciel d'administration automatisée de tests se traduit dans la diversité des types d'items. Un même test peut être composé de cinq types différents d'items, cinq étant le maximum qu'offre le système.

La troisième caractéristique importante du logiciel réfère à la possibilité offerte à l'utilisateur d'inclure dans une banque des items dont la coapparition dans un même test est jugée inacceptable. Lors de la génération d'un test, le choix d'un de ces items empêche automatiquement la présence des autres.

La quatrième caractéristique du logiciel, propre à la situation d'évaluation interactive, réside dans la souplesse offerte à l'étudiant lors de la passation d'un test. L'étudiant, en effet, répond à tous les items d'un test avant de voir ses réponses corrigées; il peut même revenir sur des items particuliers ou encore réviser son test avant d'en demander la correction. Le "feed-back" à chacune des questions n'est fourni qu'au terme de l'activité d'évaluation, au moment où l'étudiant décide d'y mettre fin.

RÉSULTATS DE L'ÉVALUATION

Les trois mois d'observation-participante nous ont conduits dans l'ensemble à une évaluation positive du logiciel d'administration automatisée de tests, compte tenu des cinq critères d'évaluation retenus.

Aucune défectuosité technique majeure n'est venue entraver le déroulement; cela est dû, croyons-nous, en grande partie à la période intense d'utilisation des deux premiers mois faite à l'INRS-Éducation lors de l'initiation des enseignants. Pendant cette période, en effet, plus de 300 items ont été inscrits dans le système et de nombreuses simulations de passation de tests au terminal ont été menées. Cette pré-expérimentation en laboratoire nous a permis, en un laps de temps très court, d'exploiter au maximum les différentes fonctions du système et d'identifier très rapidement les faiblesses techniques. Cependant, le guide d'utilisation destiné à l'enseignant n'ayant été rédigé formellement qu'en juin et juillet 1984, certains problèmes techniques mineurs ont été rencontrés. Ces difficultés par contre relevaient surtout de la nouveauté de l'instrument utilisé et du manque de familiarisation de la part des enseignants et des étudiants avec les appareils informatiques. La connaissance des diverses fonctions, des possibilités et des limites du système une fois maîtrisée par les enseignants et les étudiants, l'expérimentation nous a montré que le logiciel ne recelait pas de difficultés techniques majeures.

Les principales fonctions du système ont été utilisées proportionnellement à ce que nous avions prévu. Aucune fonction n'a été négligée ou considérée inutile par les enseignants, même si certaines ont été privilégiées. Toutefois, l'expérimentation nous a permis de rendre plus limpides certaines fonctions du logiciel, celles en particulier reliées à la passation de tests; elle nous a amenés aussi à augmenter les possibilités de certaines opérations et à en ajouter de nouvelles. Nous en traiterons dans la partie suivante de notre exposé intitulée "les améliorations apportées au logiciel".

La limite de notre évaluation résulte du fait que l'équipe de conception du logiciel a été sensiblement la même que celle chargée de son évaluation. Nous en étions conscients et le questionnaire adressé aux enseignants a été pour nous un moyen pour contourner partiellement l'écueil de notre subjectivité; il a été aussi un moyen pour évaluer les deux critères concernant respectivement les avantages et les désavantages de l'utilisation du logiciel tels que perçus par les enseignants, le degré d'appropriation du logiciel ressenti par les enseignants.

Les deux enseignants qui ont expérimenté le logiciel d'administration automatisée de tests ont été unanimes à reconnaître la valeur de cet instrument dans le cadre de l'évaluation formative des apprentissages. Pour eux, en effet, cet outil a rendu possible le développement d'habitudes d'auto-évaluation chez les étudiants; il a de plus initié progressivement les étudiants à une lecture attentive des questions posées et à la rigueur des réponses données.

Le plus grand avantage reconnu par les utilisateurs a été la souplesse inhérente au logiciel pour la production de tests variés respectant l'importance accordée à l'une ou l'autre des sections de la matière enseignée. En effet, les deux enseignants ont eu constamment recours, lors de la construction des tests, à la possibilité de l'échantillon stratifié d'items, la possibilité de la génération des items au hasard pour l'ensemble de la banque n'étant à peu près pas utilisée, sinon pour

certains tests dits de révision. Les raisons évoquées pour expliquer leur nette préférence pour l'échantillon stratifié d'items sont les suivantes. Premièrement, l'importance d'une catégorie (ou groupe) d'items n'est pas directement reliée au nombre d'items disponibles dans la banque pour cette catégorie. En second lieu, l'importance accordée à une catégorie d'items peut varier presque à l'infini sans l'obligation d'ajouter ou de retrancher des items dans la banque. Troisièmement, la prise en compte par le système d'items jugés équivalents et enfin, la qualité de la représentativité de l'ensemble des items d'un test font du logiciel, à leur avis, un outil efficace pour élaborer des tests d'évaluation, assurer la passation des tests ainsi que la correction.

En ce qui concerne le critère relatif au degré d'appropriation par l'enseignant du logiciel d'administration automatisée de tests, nous pouvons conclure qu'il est directement relié à la compréhension du système et particulièrement à l'opération de construction de la banque. Si nous nous référons tant aux avis des deux utilisateurs qu'aux rapports des deux observatrices de l'INRS-Éducation, la création d'items est l'opération la plus exigeante, car elle nécessite une réflexion soutenue concernant la formulation particulière des énoncés et des paramètres. Des questions comme, par exemple, celles concernant l'exactitude des réponses attendues, la latitude permise lors de la correction ou encore le choix du type d'items en fonction de l'énoncé sont autant d'interrogations que l'utilisateur doit avoir en tête au moment où il construit une banque. De plus, l'impossibilité dans certains cas de prévoir toutes les réponses valables lors de l'élaboration des clés de correction oblige l'enseignant à "se soumettre à de fréquentes modifications essentielles à l'amélioration de sa banque"[1].

---

1. Brouard, J., Observation participante de l'utilisation dans une classe du logiciel "Construction d'une banque d'items et passation de tests au terminal" à l'École secondaire Marie-de-l'Incarnation. INRS-Éducation, R-172, 1984.

Le haut degré d'appropriation du logiciel manifesté par les deux enseignants pendant l'expérimentation a été proportionnel à la qualité de leur travail professionnel. Si le système possède un degré de transparence qui en facilite l'appropriation, il nous faut admettre, de concert avec les deux enseignants impliqués dans l'expérience et les deux observatrices, qu'il est nécessaire que l'utilisateur accorde "un bon temps de préparation axé sur le contenu des questions, des réponses et des commentaires ajoutés aux réponses... C'est leur adéquation à la matière et à la clientèle, leur variété, leur précision, même leur nouveauté... qui feront de ce logiciel un instrument"[2] favorisant l'évaluation formative des apprentissages.

Quant aux réactions des étudiants face à l'utilisation du logiciel, nous disposons pour les évaluer des commentaires des observatrices et des enseignants ainsi que des avis des étudiants formulés soit lors de la plénière destinée à recueillir leurs opinions, soit sur la partie "appréciation" de la grille d'observations (cf. Annexe I). Ce critère est évidemment retenu en fonction de la situation interactive d'évaluation, c'est-à-dire en regard des réactions d'étudiants pendant ou après une session d'évaluation.

Dans l'ensemble, les réactions des étudiants ont été très favorables, même après la période d'excitation qu'occasionne actuellement l'installation d'un ordinateur personnel dans une école qui n'en possède pas encore. Les raisons évoquées par les étudiants et explicatives de leur satisfaction face à l'utilisation du logiciel donnent à réfléchir. Pour une très grande majorité d'entre eux, la "passation régulière de tests au terminal leur permet de mieux prendre conscience de ce qu'ils savent ou ne savent pas d'un chapitre ou d'une notion en particulier"[3].

---

2. Hébert, M.M., Observation-participante de l'utilisation dans une classe du logiciel "Construction d'une banque d'items et passation de tests au terminal" à l'École secondaire des Ursulines de Québec. INRS-Éducation, R-173, 1984.
3. Brouard, J., op. cit.

Ce genre d'évaluation est l'occasion d'approfondir une matière de "façon moins ennuyeuse que l'écoute ou la lecture"[4]; ce contexte interactif d'évaluation a le mérite, selon les étudiants des deux classes expérimentales, de favoriser l'apprentissage de notions ou leur consolidation. Et bien que "la sévérité de correction soit parfois considérée comme un désagrément, les élèves reconnaissent que cela peut les aider grandement dans leur maîtrise de la matière et de l'orthographe en particulier"[5].

Les résultats de l'évaluation en fonction des cinq critères que nous avions retenus confirment, à notre avis, la pertinence de l'utilisation par l'enseignant et ses élèves d'un système conçu pour l'administration automatisée de tests. Compte tenu de ces résultats, l'équipe de recherche chargée de la conception et de l'évaluation du logiciel a procédé au terme de la période expérimentale à la finalisation de l'instrument.

VERSION FINALE DU LOGICIEL

Nous soulignons immédiatement que le logiciel d'administration automatisée de tests est connu, depuis juin 1984 sous le nom de "Rétroaction". De plus, un guide d'utilisation[6] est disponible depuis la fin de juillet 1984, guide dont le but est de faciliter la familiarisation des futurs utilisateurs et où le lecteur peut retrouver rapidement, dès les toutes premières pages, une description claire et rapide du logiciel. Nous nous limiterons ici à souligner les principales améliorations qui ont été apportées au logiciel à la lumière de l'expérimentation.

À la demande expresse des enseignants (cela correspondait d'ailleurs à un voeu de l'équipe de recherche), une dimension nouvelle a été

---

4. Idem.
5. Idem.
6. Rhéaume, D., Brouard, J., Guide d'utilisation du logiciel éducatif "Rétroaction". INRS-Éducation, juillet 1984.

apportée à la génération des tests. Dans sa phase expérimentale, le logiciel offrait deux façons de générer un test: le mode général qui correspondait à un échantillon d'items choisis au hasard à l'intérieur d'une banque ou encore le mode stratifié basé sur l'échantillon stratifié d'items, la stratification ayant trait aux catégories d'items ou groupes d'items. La version finale du logiciel offre un troisième mode de génération de tests, un mode fixe où l'enseignant peut choisir lui-même les items devant faire partie d'un test.

La place accordée au feed-back a été augmentée. Alors que la version expérimentale n'y accordait que 30 caractères, la version finale offre une possibilité de feed-back pouvant atteindre jusqu'à 5 lignes de 40 caractères dans le cas d'items qui occupent moins d'espace que le maximum prévu. Cette modification est capitale si l'on pense que l'espace réservé au feed-back représente la latitude attribuée à l'enseignant pour fournir, en plus de la bonne réponse, des éléments de contenu ou encore des rappels à des notions antérieures, des références, des explications supplémentaires, etc. Nous notons que la nature du feed-back ne dépend pas de la nature de la réponse donnée par l'élève. C'est plutôt la rétroaction présentée à l'étudiant lorsqu'il passe un test au terminal: le feed-back pour un item est unique et c'est à l'enseignant de l'exploiter de la meilleure façon possible au moment où il construit sa banque d'items. Pour chaque item, le feed-back devrait être le lieu privilégié pour provoquer l'apprentissage.

Une autre amélioration importante apportée au logiciel concerne les deux types de questions exigeant une réponse plus ou moins brève. Dans la version expérimentale, l'utilisateur pouvait uniquement avoir accès à trois clés de correction différentes; dans la version finale, pour augmenter le degré de souplesse de la correction ou encore pour doter le système d'une plus grande flexibilité au niveau de l'énoncé même de l'item, le nombre maximum de clés de correction a été porté à huit si l'espace-mémoire le permet. La multiplicité des clés permettra à l'enseignant de prévoir, entre autre, comme bonnes réponses, des termes ou

expressions considérés comme synonymes. La restriction du nombre de clés disponibles était en partie responsable de certains problèmes lors de la passation des tests; des étudiants se sont vus refuser une bonne réponse parce que le mot qu'ils avaient utilisé, même s'il était juste, n'avait pu être introduit dans les clés disponibles.

Plusieurs modifications mineures ont aussi contribué à l'amélioration de la version finale du logiciel. Pour n'en citer que quelques-unes, soulignons la présence d'un traitement de texte plus raffiné lors de l'entrée des items ou de la présentation générale du test, le changement de certaines instructions en termes plus précis et plus clairs, une lisibilité accrue des écrans en particulier lors de la révision du test par l'étudiant, la fusion de certaines fonctions par souci de simplicité, etc. De plus, des ajouts, comme l'intégration d'un compteur de caractères disponibles lors de l'entrée d'un item, ont été incorporés au logiciel à la demande des enseignants.

Les modifications apportées au logiciel nous ont été suggérées par l'expérimentation mais elles n'ont pu faire l'objet d'une nouvelle expérimentation en classe.

AVENIR IMMÉDIAT DU LOGICIEL

Avant de terminer cet exposé, nous tenons à noter que l'expérimentation du logiciel en milieu scolaire a créé un intérêt manifeste chez les enseignants et la direction des deux écoles impliquées dans le processus d'évaluation. À l'école des Ursulines de Québec, la direction de l'institution a réservé deux journées pédagogiques complètes pour sensibiliser ses enseignants aux applications de l'ordinateur personnel; le logiciel d'administration automatisée de tests a servi, à cet égard, d'exemple pour illustrer "une façon heureuse d'introduire l'ordinateur

en milieu scolaire"[7]. Pour leur part, plusieurs enseignants ainsi que la direction de l'École Marie-de-l'Incarnation ont manifesté le désir d'obtenir une version "Commodore" du logiciel, convaincus qu'ils étaient des avantages du système et de sa relative simplicité.

Nous ne pouvons pas présumer de l'avenir du logiciel. Cependant, il importe de souligner l'entente conclue entre l'INRS-Éducation et la Télé-Université concernant la diffusion de "Rétroaction". Avant la fin de 1984, la Télé-Université l'incorporera au matériel requis pour le cours "Ordinateur environnement éducatif" sous la responsabilité du professeur Louise Sauvé et auquel devraient s'inscrire plus d'un millier d'enseignants. L'entente prévoit que le logiciel sera disponible à prix réduit pour toute personne qui en fera la demande à la Télé-Université.

---

7. Idem.